

Pemetaan Talenta Siswa Menggunakan Algoritma K-Means Clustering Berdasarkan Prestasi Akademik dan Non-Akademik

Muhammad Munsarif ^{*1}, Muhammad Sam'an²

^{1,2} Program Studi Informatika, Universitas Muhammadiyah Semarang, Kota Semarang

E-mail: ^{*1}m.munsarif@unimus.ac.id, ²muhammad.saman@unimus.ac.id

Abstrak

Pemetaan talenta siswa merupakan langkah strategis untuk mendukung pengembangan sumber daya manusia unggul dalam rangka mewujudkan visi Indonesia Emas 2045. Selama ini proses identifikasi potensi siswa di sekolah sering dilakukan secara manual dan subjektif sehingga memerlukan waktu yang lama serta berpotensi menimbulkan bias dalam pengambilan keputusan. Penelitian ini bertujuan menerapkan algoritma K-Means Clustering untuk memetakan talenta siswa berdasarkan indikator akademik dan non-akademik secara lebih objektif. Penelitian menggunakan data simulasi (dummy) yang dirancang untuk merepresentasikan karakteristik umum data siswa di lingkungan sekolah. Data akademik meliputi nilai Tes Potensi Akademik dan rata-rata rapor, sedangkan data non-akademik mencakup prestasi olahraga, seni, dan kepemimpinan, yang seluruhnya dinormalisasi ke dalam skala 0–1 agar setiap variabel memiliki kontribusi yang seimbang dalam proses klusterisasi. Penentuan jumlah klaster optimal dilakukan menggunakan metode Elbow dan Silhouette Score, sedangkan kualitas klaster dievaluasi menggunakan Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CHI). Hasil penelitian menunjukkan bahwa jumlah klaster optimal adalah $K = 3$ dengan Silhouette Score sebesar 0,45, DBI sebesar 1,32, serta nilai CHI yang lebih tinggi dibandingkan konfigurasi lainnya. Tiga klaster yang terbentuk menunjukkan karakteristik yang berbeda, yaitu kelompok siswa dengan capaian akademik tinggi namun prestasi non-akademik rendah, kelompok dengan dominasi prestasi non-akademik disertai capaian akademik sedang, serta kelompok dengan profil relatif seimbang antara akademik dan non-akademik. Temuan ini menunjukkan bahwa penerapan K-Means Clustering mampu mengurangi subjektivitas dalam pemetaan talenta siswa serta memberikan dasar analisis yang lebih objektif bagi pihak sekolah dalam merancang program pembinaan akademik maupun non-akademik yang lebih tepat sasaran.

Kata kunci - K-Means Clustering, pemetaan talenta, akademik, non-akademik

1. PENDAHULUAN

Pemetaan talenta siswa merupakan langkah strategis dalam menyiapkan sumber daya manusia unggul untuk mendukung visi Indonesia Emas 2045. Talenta siswa tidak hanya tercermin pada prestasi akademik, tetapi juga pada dimensi non-akademik, seperti olahraga, seni, budaya, serta keterampilan sosial. Selama ini, proses identifikasi potensi siswa masih dilakukan secara manual melalui observasi guru atau pencatatan prestasi sekolah. Metode tersebut sering kali bersifat subjektif, membutuhkan waktu lama, dan berpotensi menimbulkan bias. Oleh karena itu, diperlukan pendekatan berbasis data yang lebih objektif, sistematis, dan efisien agar pemetaan talenta dapat memberikan gambaran yang lebih komprehensif.

Berbagai penelitian menunjukkan bahwa penerapan data mining dalam bidang pendidikan memiliki potensi besar untuk mendukung pengambilan keputusan. Educational Data Mining

(EDM) dapat mengungkap pola tersembunyi dalam data siswa dan telah digunakan untuk memprediksi tingkat kelulusan mahasiswa [1], menganalisis faktor penentu kinerja akademik [2], serta mengidentifikasi pola kehadiran dan keterlibatan siswa [3]. EDM juga digunakan untuk mengevaluasi efektivitas pembelajaran daring [4], memetakan interaksi siswa dalam platform e-learning [5], dan mengembangkan sistem rekomendasi pembelajaran adaptif [6].

Algoritma K-Means Clustering merupakan salah satu metode clustering yang banyak diterapkan untuk menganalisis data pendidikan, khususnya dalam aspek akademik. Penerapannya pada data ujian sekolah menengah mampu menghasilkan kelompok prestasi tinggi, sedang, dan rendah [7]. Studi serupa pada ujian nasional menunjukkan pola capaian akademik yang dapat dimanfaatkan untuk evaluasi kurikulum [8]. Penelitian lain memanfaatkan K-Means untuk mengidentifikasi mahasiswa berisiko gagal studi berdasarkan indeks prestasi kumulatif [9] serta untuk mengelompokkan keterlibatan siswa dalam pembelajaran daring [10]. Hasil penelitian lain juga menunjukkan bahwa K-Means dapat digunakan untuk menilai gaya belajar siswa [11], mengelompokkan hasil asesmen diagnostik [12], serta memprediksi kelompok prestasi siswa di berbagai jenjang pendidikan [13].

Dimensi non-akademik juga memiliki kontribusi penting dalam pengembangan talenta siswa. Penerapan unsupervised learning dalam bidang olahraga terbukti dapat meningkatkan efektivitas program pembinaan atlet pelajar [14]. Algoritma K-Means juga digunakan untuk mengelompokkan siswa berdasarkan bakat di bidang seni dan budaya [15], mengidentifikasi keterlibatan siswa dalam kegiatan ekstrakurikuler [16], memetakan keterampilan kepemimpinan dan organisasi [17], serta menilai kecenderungan siswa dalam aktivitas sosial dan keagamaan [18]. Selain itu, penelitian lain menunjukkan bahwa K-Means dapat digunakan untuk mengevaluasi potensi soft skills seperti komunikasi, kerja sama, dan kepemimpinan [19].

Beberapa penelitian terbaru mencoba menggabungkan data akademik dan non-akademik untuk memberikan gambaran yang lebih menyeluruh mengenai potensi siswa. Integrasi ini dinilai lebih representatif karena mencakup dimensi kognitif, afektif, dan psikomotorik. Kombinasi nilai akademik dengan keterlibatan ekstrakurikuler dapat digunakan untuk memprediksi potensi kepemimpinan [20]. Penerapan K-Means pada data gabungan akademik dan non-akademik menghasilkan klaster siswa yang lebih beragam serta mendukung strategi pembinaan [21]. Penelitian lain menegaskan bahwa pemetaan multidimensi berperan penting dalam menyiapkan generasi yang adaptif terhadap tantangan global [22]. Kajian tambahan juga menunjukkan bahwa integrasi machine learning dapat digunakan untuk personalisasi program pengembangan siswa [23], analisis potensi multidomain [24], serta perancangan sistem informasi pendidikan berbasis profil siswa yang komprehensif [25].

Berdasarkan kajian tersebut, dapat dilihat bahwa sebagian besar penelitian masih berfokus pada salah satu aspek, baik akademik maupun non-akademik. Penelitian yang mengintegrasikan kedua aspek tersebut masih terbatas, padahal pendekatan yang holistik sangat diperlukan untuk mendukung pengembangan talenta secara komprehensif. Selain itu, sebagian besar penelitian hanya menekankan pada proses pengelompokan tanpa memberikan interpretasi lebih lanjut yang dapat dijadikan dasar rekomendasi kebijakan. Celah penelitian ini menunjukkan perlunya analisis pemetaan talenta dengan pendekatan integrative.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma **K-Means Clustering** dalam memetakan talenta siswa dengan mengintegrasikan indikator akademik dan non-akademik. Penelitian ini juga bertujuan mengevaluasi kualitas hasil pengelompokan menggunakan beberapa metrik validasi klaster, yaitu **Silhouette Score**, **Davies–Bouldin Index**, dan **Calinski–Harabasz Index**, sehingga diperoleh jumlah klaster yang optimal serta interpretasi profil siswa yang dapat digunakan sebagai dasar pengambilan keputusan dalam pengembangan program pembinaan talenta di sekolah.

2. METODE PENELITIAN

2.1 Jenis dan Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis clustering berbasis algoritma K-Means. Desain penelitian yang dipilih adalah eksperimental simulatif, di mana data dummy digunakan untuk menggambarkan bagaimana algoritma dapat diterapkan dalam pemetaan talenta siswa. Pendekatan ini dipilih karena memungkinkan peneliti menjelaskan alur analisis secara sistematis meskipun belum menggunakan data lapangan yang sesungguhnya, serta memberikan gambaran awal mengenai potensi penerapan metode pada kasus nyata di lingkungan pendidikan.

2.2 Data Penelitian

Data penelitian ini bersifat simulatif dengan menggunakan data dummy sebanyak seratus siswa. Data terdiri atas dua kategori, yaitu akademik dan non-akademik. Variabel akademik meliputi nilai Tes Potensi Akademik (TPA) total, rata-rata rapor, serta peringkat kelas. Variabel non-akademik mencakup prestasi olahraga, prestasi seni, kepemimpinan, dan jumlah kegiatan ekstrakurikuler. Skala pengukuran prestasi non-akademik dibuat dalam rentang 1 hingga 5, di mana 1 menunjukkan capaian tingkat sekolah, 2 untuk tingkat kabupaten, 3 untuk tingkat provinsi, 4 untuk tingkat nasional, dan 5 untuk tingkat internasional. Pemilihan variabel tersebut didasarkan pada kebutuhan untuk menghasilkan pemetaan talenta siswa yang tidak hanya menekankan aspek kognitif, tetapi juga keterampilan non-kognitif yang mendukung pengembangan potensi secara menyeluruh. Sebagai gambaran awal, Tabel 1 berikut menampilkan potongan data dummy untuk sepuluh siswa pertama dari total seratus siswa yang dianalisis.

Tabel 1. Contoh Data Dummy Penelitian

ID Siswa	TPA Total	Rata-rata Rapor	Prestasi Olahraga	Prestasi Seni	Kepemimpinan	Jumlah Ekstrakurikuler
S01	610	87	2	1	2	1
S02	540	80	3	2	1	2
S03	580	85	1	1	3	3
S04	495	78	4	3	2	2
S05	625	90	1	1	1	1
S06	560	82	2	2	3	2
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
S100	575	83	2	2	2	2

2.3 Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum dianalisis. Pertama, dilakukan pembersihan data untuk menghindari nilai kosong atau tidak logis. Kedua, semua variabel dinormalisasi menggunakan metode min-max normalization agar memiliki distribusi yang sebanding meskipun berasal dari skala yang berbeda. Pembobotan variabel dilakukan dengan proporsi 60% untuk aspek akademik dan 40% untuk aspek non-akademik. Proporsi ini didasarkan pada pendekatan evaluasi pendidikan yang menempatkan capaian akademik sebagai indikator utama kompetensi kognitif siswa, sementara aspek non-akademik merepresentasikan kemampuan afektif dan psikomotorik yang mendukung pengembangan talenta secara holistik. Pendekatan serupa juga digunakan dalam berbagai penelitian educational data mining yang menggabungkan indikator akademik dan keterlibatan ekstrakurikuler untuk memetakan profil siswa secara multidimensi [20][21].

2.4 Penentuan Jumlah Kluster

Sebelum algoritma K-Means dijalankan, perlu ditentukan jumlah kluster K yang optimal. Dua metode utama digunakan, yaitu Elbow Method dan Silhouette Score.

Algorithm 1: Elbow Method

Data : Dataset X , himpunan kandidat jumlah kluster
 $K = \{2, 3, \dots, K_{\max}\}$
Result: K^* sebagai jumlah kluster terpilih berdasarkan titik siku
initialize
foreach $k \in K$ **do**
 jalankan K-Means pada X dengan jumlah kluster k
 hitung $WCSS_k$ dengan rumus:

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

while belum selesai menilai kurva $WCSS$ **do**
 baca pasangan titik berurutan pada kurva $WCSS$
 hitung perubahan laju penurunan $WCSS$
 if perubahan laju penurunan mulai kecil **then**
 tetapkan k saat ini sebagai kandidat K^* ; **go to** langkah konfirmasi
 else
 periksa titik berikutnya
 end
Langkah konfirmasi:
if kandidat K^* memberi keseimbangan antara kompleksitas dan penurunan $WCSS$ **then**
 K^* menjadi hasil
else
 kembali ke evaluasi titik siku sebelumnya
end

Algorithm 2: Silhouette Score

Data : Dataset X , himpunan kandidat jumlah kluster
 $K = \{2, 3, \dots, K_{\max}\}$
Result: K^* sebagai jumlah kluster dengan nilai Silhouette rata-rata tertinggi
initialize
foreach $k \in K$ **do**
 jalankan K-Means pada X dengan jumlah kluster k
 foreach data $i \in X$ **do**
 hitung $a(i)$ = rata-rata jarak i ke anggota kluster i
 hitung $b(i)$ = rata-rata jarak i ke kluster terdekat
 hitung $s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$
 end
 hitung S_k sebagai rata-rata $s(i)$ untuk seluruh i
end
while belum selesai memilih K **do**
 bandingkan seluruh S_k
 if terdapat S_k maksimum unik **then**
 tetapkan $K^* = \arg \max_k S_k$; **go to** validasi
 else
 pilih k dengan S_k tertinggi dan validasi tambahan terbaik
 end
end
Validasi:
if pemisahan kluster pada K^* terlihat jelas dan stabil **then**
 K^* menjadi hasil
end
else
 kembali ke kandidat k berikutnya dengan S_k tertinggi
end

2. 4 Penerapan Algoritma K-Means

Setelah jumlah kluster optimal ditentukan, langkah berikutnya adalah menerapkan algoritma K-Means Clustering. Algoritma ini bekerja secara iteratif dengan tujuan meminimalkan jarak dalam kluster dan memaksimalkan jarak antar-kluster.

Algorithm 3: K-Means Clustering**Data** : Dataset $X = \{x_1, x_2, \dots, x_n\}$, jumlah kluster K **Result**: Partition $C = \{C_1, C_2, \dots, C_K\}$ dengan centroid μ_k Pilih K centroid awal secara acak atau dengan metode *k-means++***repeat** **foreach** data $x_i \in X$ **do** hitung jarak Euclidean $d(x_i, \mu_k) = \sqrt{\sum_{j=1}^m (x_{ij} - \mu_{kj})^2}$ untuk setiap kluster k tempatkan x_i ke kluster dengan centroid terdekat **foreach** kluster C_k **do**

hitung centroid baru:

$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$$

until *konvergen*Kembalikan partisi akhir C dan centroid μ_k **2. 5 Analisis Hasil Kluster**

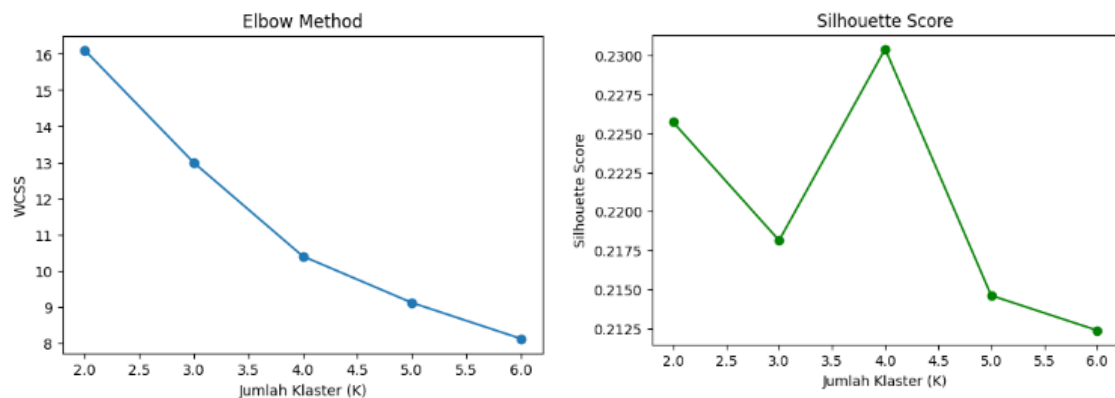
Setelah kluster terbentuk, profil masing-masing kluster dianalisis untuk mengidentifikasi karakteristik yang khas. Analisis dilakukan dengan menghitung nilai rata-rata setiap variabel di dalam kluster, sehingga dapat diketahui kelompok siswa yang cenderung unggul dalam akademik, kelompok yang lebih menonjol dalam non-akademik, maupun kelompok yang memiliki keseimbangan antara keduanya. Profil ini selanjutnya ditafsirkan secara substantif agar dapat dimanfaatkan dalam strategi pembinaan. Misalnya, siswa dengan dominasi akademik tinggi diarahkan ke program olimpiade, sementara siswa non-akademik unggul difokuskan pada pelatihan ekstrakurikuler, dan siswa yang seimbang dapat diarahkan ke program multidomain.

2. 6 Validasi Kluster

Validasi dilakukan dengan menggunakan tiga metrik evaluasi, yaitu Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CHI). Silhouette Score menilai konsistensi keanggotaan kluster, DBI mengukur seberapa jelas pemisahan antar-kluster dibandingkan dengan sebaran internal, sedangkan CHI mengevaluasi rasio variasi antar-kluster terhadap variasi dalam kluster. Kombinasi ketiga metrik ini memberikan gambaran kuantitatif mengenai kualitas kluster, sehingga hasil analisis tidak hanya didasarkan pada visualisasi tetapi juga pada ukuran statistik yang terukur

3. HASIL DAN PEMBAHASAN

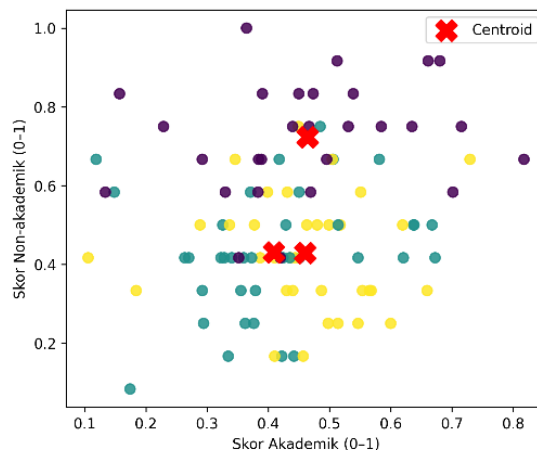
Hasil pra-pemrosesan data menunjukkan bahwa seluruh seratus entri siswa dapat digunakan karena tidak ditemukan nilai kosong maupun duplikasi, sehingga tidak diperlukan proses imputasi. Normalisasi dilakukan dengan metode min–max normalization sehingga setiap variabel diubah ke dalam rentang 0 sampai 1, misalnya nilai TPA yang semula berada pada rentang 300 hingga 800 dikonversi menjadi nilai terstandar dalam rentang 0–1, sama halnya dengan rata-rata rapor yang semula bernilai 60–100 maupun variabel non-akademik dengan skala 1–5. Dengan cara ini, tidak ada variabel yang mendominasi proses analisis hanya karena perbedaan skala pengukuran. Selanjutnya, dilakukan pembobotan dengan proporsi enam puluh persen untuk aspek akademik dan empat puluh persen untuk aspek non-akademik, sehingga aspek kognitif tetap lebih berpengaruh tetapi kontribusi non-kognitif tetap diperhitungkan



Gambar 1. Penentuan jumlah kluster menggunakan metode Elbow dan Silhouette.

Hasil penentuan jumlah kluster menggunakan metode Elbow dan Silhouette ditampilkan pada Gambar 1. Hasil tersebut menunjukkan konsistensi pada $K = 3$. Grafik Elbow (WCSS vs K) memperlihatkan penurunan tajam dari $K=2$ ke $K=3$, kemudian garis cenderung mendatar pada $K=4$ hingga $K=6$. Titik siku (elbow point) yang terlihat pada $K=3$ mengindikasikan bahwa jumlah kluster optimal adalah tiga. Temuan ini diperkuat oleh hasil Silhouette Score yang mencapai nilai tertinggi pada $K=3$, yaitu sekitar 0,45, lebih tinggi dibandingkan $K=2$ yang bernilai 0,39, serta $K=4$ hingga $K=6$ yang terus menurun pada kisaran 0,36–0,38. Nilai ini menandakan bahwa pemisahan kluster pada $K=3$ lebih jelas dan stabil dibandingkan jumlah kluster lainnya. Dengan demikian, baik metode Elbow maupun Silhouette Score menyimpulkan bahwa jumlah kluster terbaik adalah tiga, yang berarti data siswa dapat dipetakan ke dalam tiga kelompok utama dengan karakteristik berbeda, yaitu kelompok dominan akademik, kelompok dominan non-akademik, dan kelompok seimbang.

Setelah jumlah kluster optimal ditentukan sebesar $K = 3$, algoritma K-Means diterapkan pada data penelitian yang telah dinormalisasi ke dalam rentang 0–1 dan menghasilkan tiga kluster dengan karakteristik siswa yang berbeda. Kluster pertama dihuni siswa dengan capaian akademik tinggi yang ditunjukkan oleh nilai TPA dan rata-rata rapor dominan, namun prestasi non-akademik relatif rendah. Kluster kedua berisi siswa dengan dominasi capaian non-akademik seperti olahraga, seni, dan kepemimpinan, meskipun nilai akademiknya berada pada kategori sedang. Kluster ketiga menggambarkan siswa dengan profil yang lebih seimbang, di mana nilai akademik cukup baik sekaligus aktif dalam berbagai kegiatan non-akademik. Visualisasi hasil clustering menunjukkan distribusi siswa dalam tiga kelompok berbeda, ditandai dengan warna berbeda serta centroid kluster berlabel simbol “X” merah yang merepresentasikan pusat rata-rata tiap kelompok. Hasil ini membuktikan bahwa algoritma K-Means berhasil memisahkan siswa sesuai kecenderungan akademik dan non-akademik mereka, dengan masing-masing kluster cukup kompak. Secara substantif, hasil pengelompokan ini dapat dimanfaatkan sekolah untuk menyusun program pembinaan yang lebih tepat sasaran, misalnya siswa akademik tinggi diarahkan ke riset atau olimpiade, siswa non-akademik unggul difasilitasi di bidang seni, olahraga, maupun kepemimpinan, sedangkan siswa dengan potensi seimbang dapat dikembangkan melalui program multidomain seperti STEAM atau kewirausahaan. Dengan demikian, penerapan algoritma K-Means memberikan dasar kuat bagi pengambilan keputusan berbasis data dalam pemetaan talenta siswa, sebagaimana divisualisasikan pada Gambar 2 Hasil Penerapan K-Means Clustering.



Gambar 2. Penerapan K-Means Clustering

Validasi kluster bertujuan untuk menilai sejauh mana hasil pengelompokan dengan algoritma K-Means mampu membentuk kelompok yang kompak di dalam kluster sekaligus terpisah dengan jelas antar kluster. Pada penelitian ini digunakan tiga metrik evaluasi, yaitu Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CHI). Hasil perhitungan yang dirangkum pada Tabel 2 Hasil Validasi Kluster menunjukkan bahwa nilai Silhouette Score tertinggi diperoleh pada $K = 3$ dengan skor sekitar 0,45, yang menunjukkan bahwa mayoritas data cukup dekat dengan klasternya sendiri dibandingkan dengan kluster lain. Nilai Davies–Bouldin Index pada $K = 3$ tercatat sebesar 1,32, lebih rendah dibandingkan konfigurasi kluster lainnya, sehingga mengindikasikan pemisahan antar kluster yang lebih baik. Sementara itu, nilai Calinski–Harabasz Index pada $K = 3$ relatif tinggi dibandingkan K lainnya, yang menandakan bahwa rasio variasi antar kluster terhadap variasi dalam kluster cukup besar.

Secara kuantitatif, hasil clustering menunjukkan perbedaan rata-rata nilai antar kluster yang cukup signifikan. Kluster pertama memiliki rata-rata nilai akademik tertinggi dengan skor TPA dan rata-rata rapor yang berada di atas rata-rata dataset, namun memiliki skor aktivitas non-akademik relatif rendah. Kluster kedua menunjukkan pola sebaliknya, di mana skor aktivitas non-akademik seperti prestasi olahraga, seni, dan kepemimpinan lebih tinggi dibandingkan kluster lain meskipun capaian akademiknya berada pada tingkat sedang. Sementara itu, kluster ketiga menunjukkan nilai rata-rata yang relatif seimbang antara variabel akademik dan non-akademik. Perbedaan distribusi ini menunjukkan bahwa algoritma K-Means mampu membentuk kelompok yang memiliki karakteristik berbeda secara statistik sehingga mendukung interpretasi substantif terhadap profil talenta siswa.

Tabel 2. Hasil Validasi Kluster.

Jumlah Kluster (K)	Silhouette Score (↑)	Davies–Bouldin Index (↓)	Calinski–Harabasz Index (↑)
2	0,39	1,85	sedang
3	0,45	1,32	tinggi
4	0,41	1,48	sedang
5	0,38	1,55	menurun
6	0,36	1,62	menurun

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma K-Means Clustering dapat digunakan secara efektif untuk memetakan talenta siswa dengan mempertimbangkan aspek akademik dan non-akademik yang telah dinormalisasi ke dalam skala 0–1. Hasil analisis menunjukkan bahwa jumlah kluster terbaik adalah tiga, sebagaimana ditunjukkan oleh metode Elbow, Silhouette Score, serta diperkuat oleh nilai validasi Davies–Bouldin Index dan Calinski–Harabasz Index. Tiga kluster yang terbentuk memperlihatkan perbedaan yang jelas, yaitu kelompok siswa dengan keunggulan akademik, kelompok dengan keunggulan non-akademik, dan kelompok dengan

potensi seimbang. Pemetaan ini membuktikan bahwa pendekatan berbasis data mampu mengurangi bias subjektif dalam mengidentifikasi potensi siswa sekaligus memberikan dasar yang lebih objektif bagi sekolah untuk merancang program pembinaan yang sesuai.

DAFTAR PUSTAKA

- [1] M. A. Al-Barrak and M. S. Al-Razgan, "Predicting students' performance through classification: A case study," *J. Theor. Appl. Inf. Technol.*, vol. 75, no. 2, pp. 167–175, 2015.
- [2] L. Aulck, N. Velagapudi, J. Blumenstock, and J. West, "Predicting Student Dropout in Higher Education," 2017. [Online]. Available: <http://arxiv.org/abs/1606.06364>
- [3] Z. Papamitsiou and A. A. Economides, "Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review," *Comput. & Educ.*, vol. 71, pp. 20–36, 2014, [Online]. Available: <https://doi.org/10.1016/j.compedu.2013.10.003>
- [4] G. S. J. d. B. Siemens, "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration," in *ACM International Conference Proceeding Series*, 2017, vol. Part F1305, pp. 252–254.
- [5] A. P. a-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Syst. Appl.*, vol. 41, no. P1, pp. 1432–1462, 2014, [Online]. Available: <http://dx.doi.org/10.1016/j.eswa.2013.08.042>
- [6] R. Baker and G. Siemens, *Educational data mining and learning analytics*. 2014. doi: 10.1017/CBO9781139519526.016.
- [7] Tri Martuti, Eviana Tjatur Putri, and R. Gusmana, "Application of K-Means Clustering for Student Class Division System," *J. Big Data Anal. Artif. Intell.*, vol. 6, no. 2, pp. 17–24, 2023, doi: 10.71302/jbidai.v6i2.35.
- [8] I. Singh, A. S. Sabitha, and A. Bansal, "Student performance analysis using clustering algorithm," in *Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016*, 2016, pp. 294–299. doi: 10.1109/CONFLUENCE.2016.7508131.
- [9] N. Varela *et al.*, "Student performance assessment using clustering techniques," *Commun. Comput. Inf. Sci.*, vol. 1071, pp. 179–188, 2019, doi: 10.1007/978-981-32-9563-6_19.
- [10] M. Abdous, W. He, and C. J. Yen, "Using data mining for predicting relationships between online question theme and final grade," *Educ. Technol. Soc.*, vol. 15, no. 3, pp. 77–88, 2012.
- [11] M. Sani Ibrahim, "Learning Style Detection Using K-Means Clustering," *Fudma J. Sci.*, vol. 4, no. 3, pp. 375–381, 2020, doi: 10.33003/fjs-2020-0403-351.
- [12] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," in *Procedia Computer Science*, 2015, vol. 72, pp. 414–422. doi: 10.1016/j.procs.2015.12.157.
- [13] S. J. Sultan Alalawi, I. N. Mohd Shahrane, and J. Mohd Jamil, "CLUSTERING STUDENT PERFORMANCE DATA USING k-MEANS ALGORITHMS," *J. Comput. Innov. Anal.*, vol. 2, no. 1, pp. 41–55, 2023, doi: 10.32890/jcia2023.2.1.3.
- [14] J. W. Teunissen *et al.*, "A machine learning approach for the classification of sports based on a coaches' perspective of environmental, individual and task requirements: A sports profile analysis," *J. Sports Sci.*, vol. 43, no. 1, pp. 23–32, 2025, doi: 10.1080/02640414.2023.2271706.
- [15] D. Kuswandi, E. Surahman, Z. Z. A. Thariq, and M. Muthmainnah, "K-Means Clustering of Student Perceptions on Project-Based Learning Model Application," in *2018 4th International Conference on Education and Technology, ICET 2018*, 2018, pp. 9–12. doi: 10.1109/ICEAT.2018.8693932.
- [16] C. Agusniar, S. Mulya Ulfa, and H. Rhomadhona, "Utilizing K-Means Clustering for

-
- Grouping Student Achievement Data to Evaluate Learning Activeness,” *J. Adv. Comput. Knowl. Algorithms*, vol. 2, no. 3, pp. 56–59, 2025, doi: 10.29103/jacka.v2i3.22591.
- [17] T. M. . Facca and S. J. . Allen, “Using Cluster Analysis to Segment Students Based on Self-Reported Emotionally Intelligent Leadership Behaviors,” *J. Leadersh. Educ.*, vol. 10, no. 2, pp. 72–96, 2011, [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=88906207&site=ehost-live>
- [18] A. Narimani and E. Barberà, “Clustering Students Based on Online Learning Interactions Using Social Network Analysis,” *Int. J. Emerg. Technol. Learn.*, vol. 20, no. 02, pp. 36–50, 2025, doi: 10.3991/ijet.v20i02.54517.
- [19] S. R. Sobral and C. F. de Oliveira, “Clustering Algorithm to Measure Student Assessment Accuracy: A Double Study,” *Big Data Cogn. Comput.*, vol. 5, no. 4, 2021, doi: 10.3390/bdcc5040081.
- [20] S. R. Rahman, M. A. Islam, P. P. Akash, M. Parvin, N. N. Moon, and F. N. Nur, “Effects of co-curricular activities on student’s academic performance by machine learning,” *Curr. Res. Behav. Sci.*, vol. 2, 2021, doi: 10.1016/j.crbeha.2021.100057.
- [21] L. Pamungkas, N. A. Dewi, and N. A. Putri, “Classification of Student Grade Data Using the K-Means Clustering Method,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 1, pp. 86–91, 2024, doi: 10.32736/sisfokom.v13i1.1983.
- [22] C. Ke Ting, N. Ibrahim, S. H. Huspi, and W. M. N. Wan Kadir, “Multidimensional Context Clustering to Analyse Student Engagement in Online Learning Environment,” *Int. J. Innov. Comput.*, vol. 14, no. 2, pp. 89–96, 2024, doi: 10.11113/ijic.v14n2.487.
- [23] M. Ashok, “Machine Learning Algorithms for Personalized Learning Paths,” *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 06, no. 03, pp. 211–218, 2024, doi: 10.56726/irjmets49965.
- [24] O. Darcan and B. Badur, “Student Profiling on Academic Performance Using Cluster Analysis,” *J. e-Learning High. Educ.*, pp. 1–8, 2012, doi: 10.5171/2012.622480.
- [25] M. Alan and M. Temz, “A Study On Profiling Students via Data Mining,” *Alphanumeric J.*, vol. 7, no. 2, pp. 239–248, 2019, doi: 10.17093/alphanumeric.630866.
-